# OCT Image Segmentation Using Neural Architecture Search and SRGAN


Omid Dehzangi, Saba Heidari Gheshlaghi, Annahita Amireskandari, Ali Rezai, Nasser Nasrabadi
Rockefeller Neuroscience Institute,
West Virginia University, Morgantown, USA

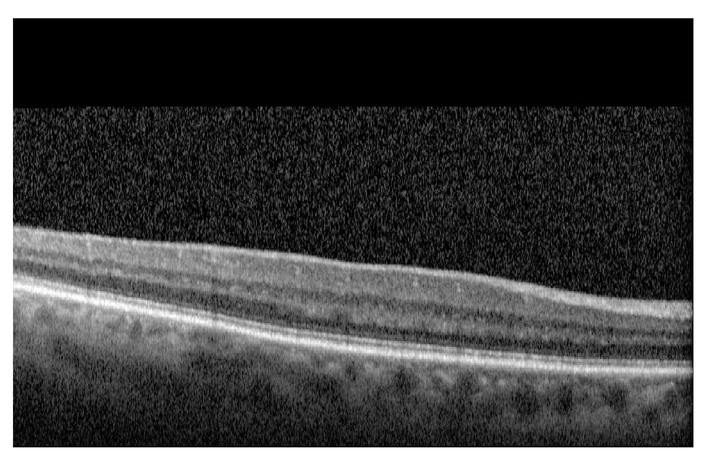

## Introduction and Motivation:

- **5.8 million people in the US are suffering from AD disease.**
- **AD is currently ranked as the sixth leading cause of death in the United States.**
- **necessary to expand our understanding of this irreversible and progressive brain disorder.**
- **Different medical procedures are used to diagnosis AD:**
  - **Position Emission Tomography (PET)**
  - **Magnetic Resonance Imaging (MRI)**
  - **Optical Coherence Tomography (OCT)?**
    - **Hypothesis: retina layer thickness change**
    - **The first step: retina layer segmentation**

## Pre-processing:

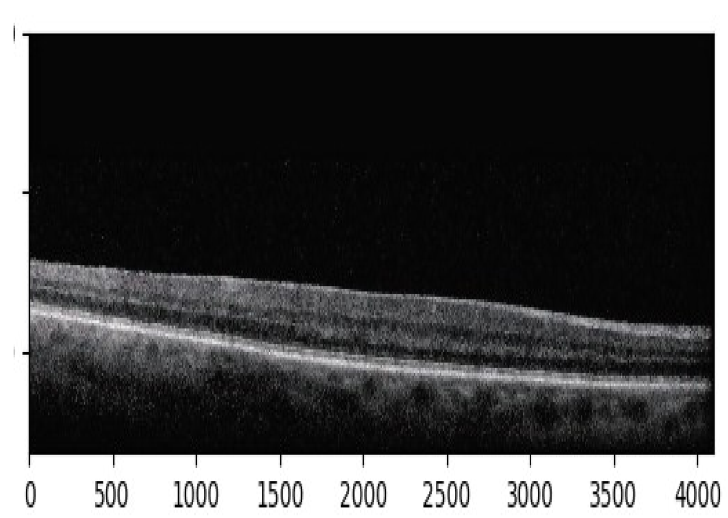
### Data Augmentation:

- **cropping the original 340×1024 OCT images to 300×300 images with an 50% overlapping.**
- **horizontal flipping applied for further data augmentation**



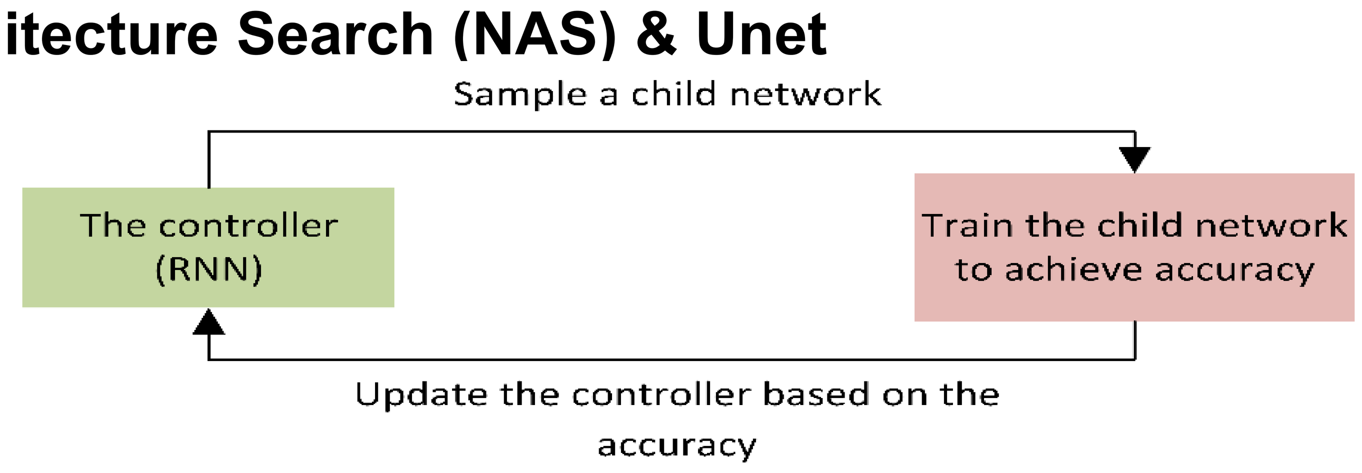
### Improving the Resolution:

- **Super Resolution GAN (SRGAN)**
- **upsamples these images by a factor of four producing HR images**
- **consists of 16 residual blocks, and the loss is calculated using the Euclidean distance between the feature maps from the network and pre-trained VGG19.**
- **This SRGAN step helps us to have higher resolution OCT images, so that the NAS network will work more efficient to find the best Unet network**




## Methodology:

### Neural Architecture Search (NAS) & Unet

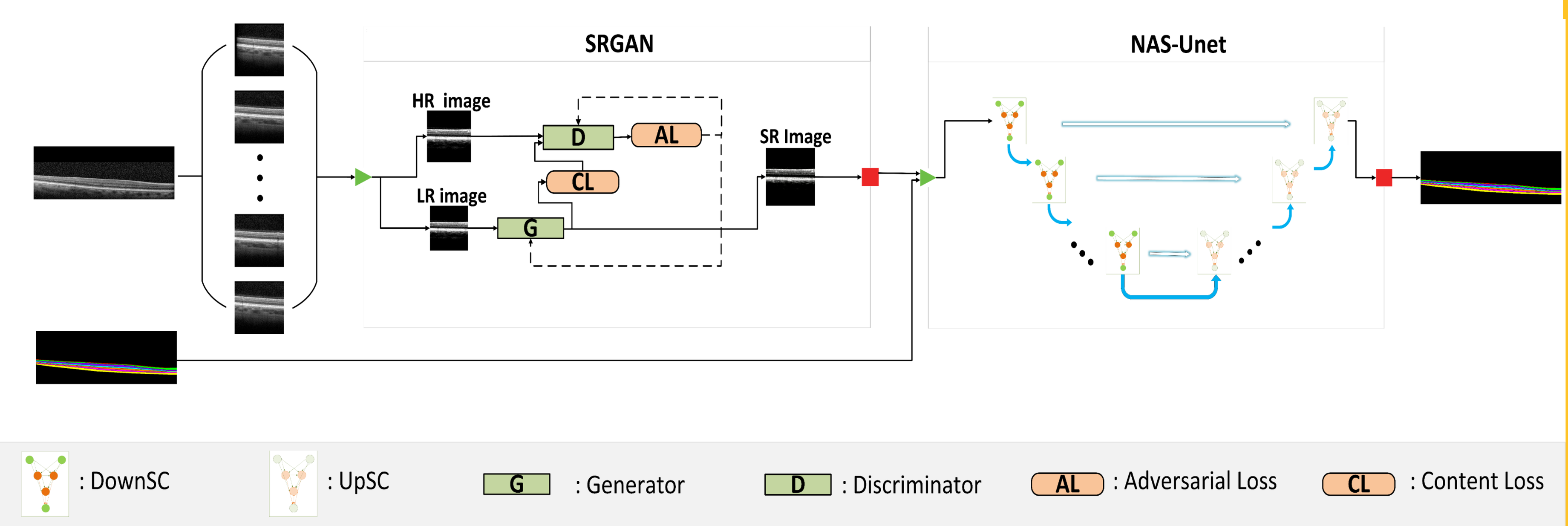



- **Finding best architects ➔ need human experts, time-consuming**
- **Neural architecture search (NAS) ➔ subfield of AutoML.**
- **NAS could find the best network architecture by maximizing the performance of the evaluation data.**
- **NAS process has categorized into three steps:**
  - **Search space**
  - **Search strategy**
  - **Performance estimation strategy**

**Unet:**

- **Famous architecture in the field of medical image semantic segmentation.**
- **A encoder-decoder (DownSampling-UpSampling) network.**
- **Encoder ➔ extracts the high-level context (semantic features)**
- **Decoder ➔ regenerates the spatial information and pixel classification results (image reconstructs)**
- **NAS-Unet algorithm  search for down- and up-sampling blocks that yield a higher accuracy on validation set.**

### The proposed architecture:

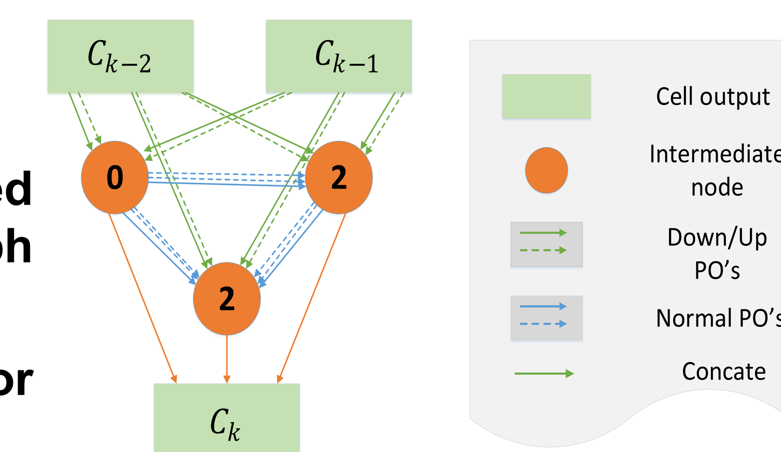



### NAS-Unet Search Space:

- **NAS requires massive computational resources to achieve good results.**
- **To tackle this problem ➔ repeatable cells and keep the backbone network fixed is introduced.**
- **In this work, NAS algorithm searches for the U-like backbone architecture. The down-sampling Cell(DownSC) and up-sampling Cell (UpSC)**

### NAS-Unet Search Strategy:

- **The search space is mostly represented by using a single directed acyclic graph (DAG).**
- **NAS will look for the best operation for each cell.**




- **idea of ProxylessNAS is used which could directly learn the architectures for large-scale tasks**
- **changes the search space from a discrete set of candidates to a continuous one.**
- **helps the algorithm to be faster and more generalized in both recurrent and convolutional architectures.**

- **two NVIDIA GeForce RTX 2080 Ti GPU having 12 GB**
- **takes about 1 day to run on the OCT scan dataset.**
- **SRGAN with kernel size of (5 ×5) and scaling factor 4.**
- **7 intermediate nodes for both DownSC and UpSC blocks.**
- **300 epochs**
- **SGD optimizer with momentum 0.95**
- **cosine learning rate in the range of 0.025 to 0.01**
- **weight decay of 0.0003**

**Various types of primitive operations (PO) used for searching the best cell architectures**

| DownSC operations | UpSC Operations | Normal Opeations |
|---|---|---|
| Average pooling | Up cweight | Identity |
| Max Pooling | Up deph conv | Cweight |
| Down Conv | Up Conv | Conv |
| Down cweight | Up dilation conv | Dilation Conv |
| Down Dilation Conv | | Depth Conv |
| Down depth Conv | | |

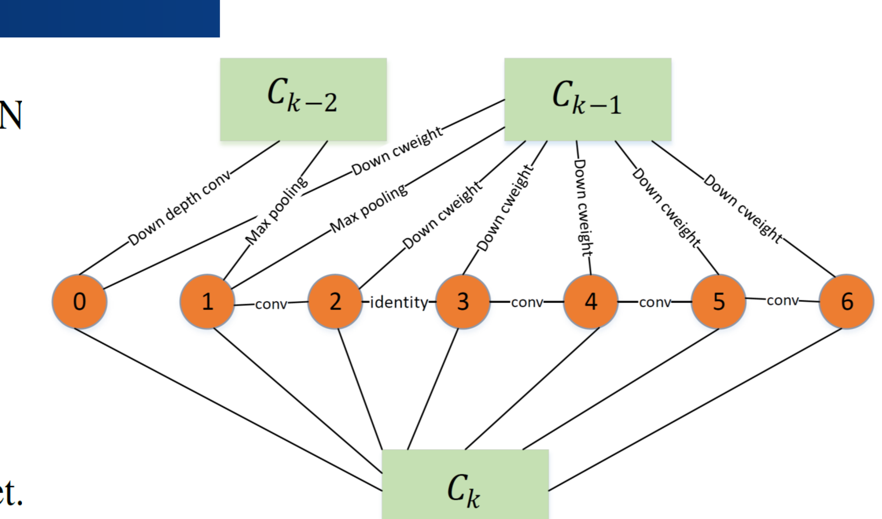
## Results:

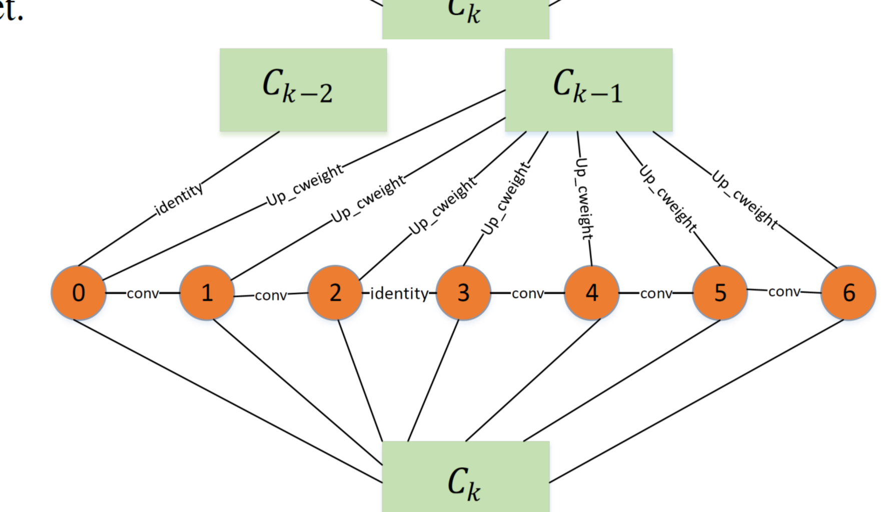
TABLE II: Comparison of NAS-Unet with/without SRGAN and human-designed Unet.

| Model | mIou | DSC | Time |
|---|---|---|---|
| Unet | 0.68 | 0.649 | **6h** |
| NAS-Unet without SRGAN | 0.945 | 0.769 | 1d-1h |
| NAS-Unet with SRGAN | **0.951** | **0.791** | 1d-6h |

TABLE III: NAS-Unet network results on our OCT dataset.

| Epoch | Pixel Accuracy | mIou | DSC |
|---|---|---|---|
| 50 | 0.453 | 0.422 | 0.438 |
| 100 | 0.602 | 0.598 | 0.486 |
| 150 | 0.835 | 0.831 | 0.632 |
| 200 | 0.892 | 0.890 | 0.707 |
| 250 | 0.926 | 0.924 | 0.771 |
| 300 | **0.952** | **0.951** | **0.791** |




## Conclusion:

- **Our proposed SRGAN NAS-Unet network was evaluated by training on our in-house OCT dataset for retina layer segmentation with high precision. This information is critical toward OCT-based AD diagnosis.**

http://www.wssplab.net/